Thursday
October 28, 1999

Part II

Department of Health and Human Services

Health Care Financing Administration

42 CFR Parts 409, 410, 411, etc.

Medicare Program; Prospective Payment System for Home Health Agencies; Proposed Rule

TABLE 9—RELATIVE CASE-MIX WEIGHTS CORRESPONDING TO HOME HEALTH RESOURCE GROUPS—Continued

HHRG group	HHRG description	Case mix weight
C3F1S3	"Clinical=High, Functional=Low, Service=High"	2.1675
C3F2S0	"Clinical=High, Functional=Mod, Service=Min"	1.1550
C3F2S1	"Clinical=High, Functional=Mod, Service=Low"	1.2389
C3F2S2	"Clinical=High, Functional=Mod, Service=Mod"	2.0674
C3F2S3	"Clinical=High, Functional=Mod, Service=High"	2.2894
C3F3S0	"Clinical=High, Functional=High, Service=Min"	1.2013
C3F3S1	"Clinical=High, Functional=High, Service=Low"	1.2852
C3F3S2	"Clinical=High, Functional=High, Service=Mod"	2.1138
C3F3S3	"Clinical=High, Functional=High, Service=High"	2.3358
C3F4S0	"Clinical=High, Functional=Max, Service=Min"	1.4357
C3F4S1	"Clinical=High, Functional=Max, Service=Low"	1.5196
C3F4S2	"Clinical=High, Functional=Max, Service=Mod"	2.3481
C3F4S3	"Clinical=High, Functional=Max, Service=High"	2.5702

4. Application of the Clinical Model Patient Classification System

The following are several illustrative examples.

Case 1

An 83-year-old woman was discharged from a hospital 2 days ago after admission for a stroke and referred for home health care. She has residual right hemiparesis and also has diabetes and hypertension. She is able to dress her upper body if clothes are laid out for her, but needs help putting on socks, nylons and sometimes slacks. She needs assistance with bathing to get in and out of the tub and uses a cane for ambulating on flat surfaces and to transfer from sitting to standing, but needs another person's assistance to go up and down stairs. She is occasionally incontinent of urine, especially at night.

Her plan of care includes—

Physical therapy: two 45-minute visits per week for 9 weeks
Occupational therapy: one 45-minute visit per week for 4 weeks
Skilled nursing: one visit per week for 2 weeks, then one visit every other week for 7 weeks
Aide: one visit twice a week for 9 weeks
Scoring: Clinical Severity=19 (for neurologic diagnosis)+8 urinary incontinence=27 high severity
Functional Status Domain=4 (for dressing)+9 (bathing)+6 (locomotion)=19 Moderate severity
Service Domain=2 (hospital discharge)+4 (therapy more than 8 hours) Moderate severity
HRG=C3F2S2

Case 2

A 73-year-old man with amyotrophic lateral sclerosis (ALS) is referred for home health care after a hospitalization for an aspiration pneumonia. Because of his inability to swallow, he had a gastrostomy tube placed during the hospitalization and now receives enteral feeding. He is dependent in all activities of daily living (ADLs).

His plan of care includes—

Skilled nursing three times a week for 9 weeks
Aide services daily for 9 weeks
Scoring
Clinical severity=19 (for neurological)+20 (for enteral feeding) High
Functional status=27 High severity
Service Domain=0 Minimum severity
HRG=C3F3S0

5. Background on Case-Mix Research Project for a National Home Health PPS

In 1996, in anticipation of the Medicare program's eventual adoption of OASIS assessment data, we began research with a sample of 90 HHAs to develop a case-mix adjustment system for use under a future national prospective payment for home health care. The project was conducted under contract to Abt Associates, Inc., of Cambridge, Mass. (Contract Number 500–96–0003/TO2). Agencies participating in the sample have collected OASIS data supplemented by approximately 50 additional assessment items on all patients newly admitted between October 1997 and April 1998 (this group of patients is called the six-month cohort) to enable comparisons among items in terms of their utility in measuring case mix. At the same time, agencies in the study collected data on every home health visit to members of the cohort. Visit information was collected on visit logs specially designed for each home health service discipline (skilled nursing, physical therapy, medical social work, etc.). The visit logs provided the fundamental measure of resource use for developing case-mix groups. This measure is the visit time, which is converted into a standardized resource cost using Bureau of Labor Statistics hourly wage data (see below for further description).

The development of case-mix groups requires identifying groups of patients with similar resource cost and similar clinical and functional characteristics. To do this, data analyses studied the statistical association between clinical and functional characteristics, as measured by the assessments, and resource cost, as measured by the standardized resource cost. In choosing patient characteristics for inclusion in the case-mix adjuster, and in arranging those characteristics into a system of groups, the system's developers gave considerable weight to the clinical diagnostic process. We sought data elements and an overall system that reflected a clinician's perspective when confronted with a patient with care needs to be assessed. We also gave considerable weight to simplicity in the system's overall structure, and thus opted for a straightforward three-dimensional approach. Under this approach, a patient's case-mix classification is found by assessing the patient on each of the three dimensions, and then combining the results from the three dimensions. Further details on the methods of the study and the resulting case-mix system follow.

Methods

Sample Selection

Agencies were recruited for the case-mix research in the spring of 1997. The sample design was intended to permit the computation of nationally representative results. Eight States (Arkansas, California, Florida, Illinois, Massachusetts, Pennsylvania, Texas, and Wisconsin) were selected to be representative of four census geographic regions: northeast, north central, south, and west. Sample selection was also intended to ensure that the four major auspices types (freestanding for-profit, freestanding voluntary/private nonprofit, hospital-based, and government) and both urban and rural agencies would be included. In addition, selection criteria included the historical practice pattern of the agencies, in order to ensure representation of agencies with relatively low, moderate, and high numbers of visits per episode in their region. When cross-classified, the four selection criteria—region (four classes), auspices (four classes), urban/rural (two classes), and practice pattern (three classes)— produced a theoretical stratification scheme consisting of 96 cells. Target sample sizes for the cells were proportional to the universe populations of the cells (for example, some of the cells had zero agencies in the universe), and totaled 90 agencies for the sample overall. To be selected, agencies had to have active Medicare certification before July 1, 1993, at least 50 Medicare patients in CY 1995, could not be participating in other HCFA demonstrations involving collection of OASIS data, and could not have been participating in the treatment group of the per-visit home health prospective payment demonstration.

Considerable effort was made to recruit and inform potential participants of the study goals and operations, and potential benefits to themselves. Potential participants were told they could expect to receive three main benefits from participation— management reports based on the data to be collected during the study, technical assistance and training on OASIS procedures, and reimbursement for data collection costs. Out of 1,797 eligible providers, approximately 290 agencies actually volunteered to participate in the study. Agencies were randomly selected from among the volunteers within each sampling cell in July 1997. Further details of the recruitment process are provided in Abt Associates, First Interim Report, July 1998 (revised December 1998).

Agency Training

The next phase of the study was training the agencies in data collection procedures. Abt Associates staff developed a Procedures Manual covering the project overview, directions on administering patient assessments using the OASIS and supplemental items (OASIS and the supplemental items were termed OASIS+, data storage and transfer procedures, and information on training techniques for agencies to use internally with their staff. Particular attention was given to item-by-item guidelines for OASIS elements, in part to ensure the reliability of the data collected for developing the case-mix adjuster. The uniform assessments afforded by OASIS were a strength of the project, because reliable data allow analysts to accurately evaluate the contribution of potential case-mix variables to a case-mix adjuster.

Additional training activities included slides and other written materials, and 2-day training sessions for participants. At least one training session was held in each of the 8 States in July and August of 1997. Training sessions were attended by 296 staff from the 90 participating agencies, and covered the meaning and intent of the OASIS and other assessment items, as well as operational procedures and data management. A significant effort was made to educate staff in methods of training and motivating their colleagues at the participating agency. After the sessions, follow-up training activities and other educational contacts were conducted by the contractor. Once the study was underway, Abt Associates continued to promote communication with the agencies, and to foster information-sharing among agencies, through activities such as conference calls, meetings, and an e-mail discussion group.

Data Resources

The two basic data sources for the study are case-mix explanatory variables from the patient assessments and a resource use variable from the visit data. Claims data comprised a third data source, and were used to verify membership in the 6-month cohort and to supply several additional potential case-mix explanatory variables for testing. All three sources of data were collected on the 6-month cohort from admission until the end of home care in the participating agency or March through April 1999, whichever came first.

OASIS data. Study agencies collected patient characteristics data using the OASIS assessment supplemented by additional assessment items at the following points: admission to home health, resumption of care following an inpatient stay, at follow up (every 57 to 62 days until discharge), upon transfer to an inpatient facility, and at discharge or death at home. The 129 patient data elements cover the following domains: patient demographics and health history, living arrangements, supportive assistance, sensory status, integumentary status, respiratory status, elimination status, neuro/emotional/behavioral status, ADLs and IADLs, medications, equipment management, emergent care use, and discharge disposition. The items supplemental to OASIS were integrated in the following OASIS domains: demographics and patient history; living arrangements; supportive assistance; integumentary status; elimination status; neuro/emotional/behavioral status; ADLs and IADLs; and medications. An additional dimension was added to the assessment data set, nutrition/hydration status, as the research literature indicates that nutritional status and the potential for dehydration are important predictors of poorer outcomes. Development of new items was beyond the scope of the project; therefore, supplemental items generally came from previously validated instruments such as the Minimum Data Set for Home Care (MDS–HC) (Morris, J. N., B. E. Fries, and D. Mehr, et al. ''A Comprehensive Clinical Assessment in Community Settings.'' November 1996a, unpublished manuscript; and Morris, J. N. The Minimum Data Set for Home Care. Presentation for ''The Key to Elderly Care in an Aging World'' in Reykjavik, Iceland, 1996b).

Visit log data. Visit information was recorded on a visit log separately tailored for each type of visit (for example, home health aide or medical social worker). The visit log consists of identifying information, starting and ending times, and a column of items for checkoff that detail the services performed during the visit and factors explaining the time spent. The checkoff items were not intended to capture information on all activities performed in the home—only those likely to significantly affect the length of the visits. The starting and ending times allow the calculation of total visit time for the key resource use measure for the study. To arrive at a standardized measure of resource use, time is weighted by the average labor cost for the discipline of the clinician making the visit.

Standardized measure of resource use. Previous research on case mix generally used a measure of resource use based on the count of visits. However, visit lengths may vary substantially, making visit counts a relatively imprecise measure of resource use. The case-mix study measured time spent on visits, rather than the number of visits themselves, to provide a more reliable measure resource use than did previous research. The mean labor cost estimate for the standardized resource use measure was based on hourly wage data from HHA respondents to the U.S. Bureau of Labor Statistics Occupational Employment Survey (OES). The survey collects wage data by occupation and industry. The Standard Industrial Classification industry category used for our estimate excludes agencies under government auspices and hospital-based agencies where workers are employed by the hospital. However, government civil service grades or hospital pay for specialized occupations may systematically depart from market wage rates. Our mean labor cost included an estimate of benefits. Following our salary equivalency estimates for therapists, the benefits were estimated exclusive of supplemental pay. The occupational category mix within each discipline (for example, registered nurses and licensed practical nurses delivering skilled nursing visits) was estimated from the OES data. For further details on the derivation of the mean labor cost used in the study, see Appendix E in Abt Associates, Inc., First Interim Report, July 1998, Revised December 1998.

Medicare claims. The Medicare claims for the 6-month cohort were linked to the patient characteristics data and visit log data to verify membership in the 6-month cohort and to provide utilization measures (for example, therapy use or institutional health care services received during the episode). The Medicare claims were also used to simulate 60-day episodes, using the from-and through-dates on the claims.

Data collection and management. The project's data management procedures were designed to support agencies in the collection and submission of consistent and reliable data on patient characteristics and service use. Participating agencies entered the patient assessment data into an electronic data file using software provided by Abt Associates or their own data systems. Data entry on site was required because this allowed a computer program to edit the data and to report any errors for correction before the data were submitted to Abt Associates. The visit logs were printed in different colors to minimize the chances for confusion. The forms were designed for optical scanning of the checkoff boxes, and the agencies forwarded the originals directly to an optical scanning contractor. The data were double entered and scanned, and the hard copy forms were sent to Abt Associates, along with the electronic data files, for cleaning. Abt processed all visit log forms received from project agencies, and generated reports for the agencies indicating the outcomes of this editing process. When agencies received the error reports and the associated hard copy logs, their responsibility was to review the problems, make any changes, and resubmit the forms.

Data preparation. The OASIS and other assessment items that had been submitted by agencies had to be merged with the records for cohort patients as defined using the claims data. Iterative matching algorithms, and intensive manual review of potential matches, were used to match assessment records to the claims patient records. Of 21,426 patients identified for the 6-month cohort from claims, 17,351 had one or more assessments that could be matched at the time Abt Associates constructed the analytic file used for case-mix system development. Visit logs on more than 750,000 visits that had been submitted by project agencies and processed by August 1998 were available for matching to claims records. Because of the occasional presence of inaccurate data in the identifying fields on the visit logs, it was necessary to protect against false matching based on incorrect visit log data. Even with an exact match on one key matching field (besides the necessary match on provider, discipline and date), it was required that the rest of the key fields be compatible. To accomplish this, a matching algorithm was developed by Abt Associates and applied to comparisons of all possible match fields. Based on the algorithm, 588,846 logged visits were matched to claims for cohort patients. The remaining logs come from visits to non-cohort Medicare patients at participating providers and visits to non-Medicare patients, inasmuch as some agencies completed logs for all of their home care patients, regardless of payor, to simplify recordkeeping procedures during the study. In addition, some of the unmatched logs likely come from an unknown number of visits to patients in the 6-month cohort whose identifying information was not sufficient to make a match at the time of file construction. (For further details of these matching procedures, see Abt Associates, Second Interim Report, August 1999.)

Analytic file construction. The project data were assembled to simulate a 60-day episode. In order to estimate resource use for each 60-day period of care, we developed certain decision rules for allocating claims and visit logs by discipline to 60-day ''windows'' of time, or episodes. Because we superimposed the 60-day episodes on the pre-existing claims stream, an episode could start and end sometime during the period covered by a claim. Many claims did not show the date of each visit; therefore, an algorithm was needed to allocate visits when a claim period fell into more than one episode. In general, the visit logs were used to make this allocation since they provided individual visit dates. If some logs were missing, the percentages of nonmissing logs falling in the claim service period before and after the episode date boundary were used to allocate visits identified on the claim to the two episodes straddled by the claim. If no logs were available, the visits from claims were allocated to the episodes in proportion to the number of days covered by the claim that fell in each 60-day episode. In episodes with missing logs, additional steps were taken to estimate the missing minutes of care that would have been measured in the missing logs. Efforts were made to use all available patient-and discipline-specific information in the imputation. Combining these procedures with a rule requiring a 60-day gap in service before a new start of care could be initiated for a cohort member resulted in a total of 31,725 payment episodes—an average of approximately 1.4 60-day episodes per cohort member with the data available at the time of file construction. After resources were calculated for all payment segments, analysis of the data revealed the presence of extreme values of mean minutes per visit by discipline within the 60-day episode. Visit lengths in episodes with extreme values (defined as the highest and lowest 0.25 percent of cases within each home health discipline) were replaced with agency-level mean visit lengths by discipline. A total of 335 episodes (1 percent) were adjusted in this manner, resulting in an insignificant change in mean total resources per 60-day episode. These allocation, imputation, and data adjustment procedures are described in detail in Abt Associates, Inc., Second Interim Report, August 1999.

Linking the Assessment Data

To complete the analytic file, the patient assessment data had to be added to the simulated episode file that contained data on visits and resource costs. To protect the reliability of the assessment data for the purpose of case-mix system development, assessments were linked to an episode in the simulation file only if the assessment was conducted within 14 days of the start of the episode.

Analytical Approach

Initial development of the case-mix model used data from 4,303 episodes pertaining primarily to the first 60-day period of care for members of the 6-month cohort who enrolled from October 1997 through December 1997. Subsequent refinement of the model occurred after the analytic file was enlarged with data accumulated later to create an augmented file. The augmented file was partitioned into a development sample and a validation sample. The development sample, consisting of 10,413 initial 60-day episodes for cohort members and 2,059 subsequent episodes, was used for the refinement phase. The development sample episodes were randomly selected from the augmented file. The remaining episodes—6,963 initial episodes and 1,331 subsequent episodes—were reserved to validate the final model.

The basic approach to case-mix development was to use the patient data and other appropriate data to identify candidate case-mix adjusters or their components, and then estimate their ability to explain variation in resource use over the course of the simulated 60-day episode. The measure of ''explanatory power'' used to evaluate the overall system and its component dimensions as development proceeded was the coefficient of determination, or R-squared.

The R-squared measures the proportion of variation in standardized resource costs that is explained by the case-mix groups. R-squared cannot be negative or greater than one. An R-squared of one would indicate that each case-mix group's average resource cost exactly predicts the individual resource cost of each episode in the case-mix group. In actual applications in social science research, an R-squared of one could be obtained only if each observation comprised its own group. The R-squared for the final home health case-mix model is .32. Based on the R-squared results, the home health case-mix system has predictive accuracy comparable to its counterparts from other payment systems. The diagnosis-related group (DRG) system used for hospital PPS has an R-squared reported in various studies in the range of .26 to .33 (Worthman, Linda G. and Shan Cretin. Review of the Literature on Diagnosis Related Groups, A RAND Note, N–2492–HCFA, Santa Monica, CA, October 1986). The Resource Utilization Groups (RUGS)–III system of 44 case mix groups used for Medicare SNF per diem prospective payment has a reported R-squared as high as .56 (Fries, B. E., D. P. Schneider, and W. J. Foley, et al., ''Refining a Case-Mix Measure for Nursing Homes: Resource Utilization Groups (RUG/II).'' *Medical Care* 32:668–685, 1994). But comparisons between the SNF and home health case-mix measures must recognize that home health resource consumption is being ''predicted'' over a 60-day period rather than on a daily basis, and that factors other than case mix may be a stronger influence on resource consumption under home health, leaving less variation to be explained by case-mix variables. Additionally, there is evidence that the RUGS–III system in actual application under the Medicare program will achieve an R-squared of less than .56 (White, A., S. Pizer, and C. White. Refining Resource Utilization Groups (RUG–III) for a National Skilled Nursing Facility System: Technical Expert Panel Briefing. October 1998).

To construct alternative case-mix groupings, preliminary regression analyses were used to investigate the relative importance of various factors explaining resource use. Then, clinical judgment was used to identify and define clinically meaningful dimensions of case mix, taking into account the results from the regressions. Alternative ways of measuring and constructing the dimensions and relating them to one another in a complete structure were explored in consultation with clinical experts. Along with clinical considerations, policy and incentive implications of alternative variables or structures were also considered—particularly the implications of alternatives for promoting improvement in health and functional status and for making the adjuster vulnerable to manipulation for profit-maximization.

Another consideration was ease of implementing the system. For example, if all of the case-mix elements were available on the OASIS assessment, then adoption of the data collection procedures necessary for PPS would already be accomplished when agencies met the OASIS requirements of the revised Conditions of Participation, pending for the quality system. Thus, the resulting case-mix groupings, and their component dimensions, were evaluated and refined interactively with clinical, policy, and administrative input.

Case-mix development work under the Abt Associates contract produced two alternative case-mix models, dubbed the ''clinical'' model and the ''diagnostic'' model. The two models had many elements in common, but the diagnostic model gave more emphasis to medical diagnosis in measuring case mix. In the diagnostic model, patients were classified into one of seven diagnosis groups based on the home health primary diagnosis from the OASIS. Further subgrouping of the basic seven groups was based on clinical, functional, and utilization-related variables. There has been controversy regarding the relative advantages and disadvantages of a diagnostically-driven model. Proponents believe it more accurately reflects the way clinicians think about patients. It may also have the potential to create more homogeneous patient groupings, providing an opportunity to develop clinical, functional, and utilization criteria customized for different diagnoses. There are several disadvantages of the diagnostically-driven model, however. One is that only a relatively few diagnostic categories (notably orthopedic, neurological, diabetes, and skin wounds/lesions) carried significant explanatory power in the analyses. This suggests that diagnostic classification beyond these few categories brings little or no additional benefit in predictive accuracy. Also, the diagnosis-based approach usually leads to a model with a higher number of end-points that may make it more complex and difficult to use. Another disadvantage is that the use of diagnostic categories is problematic when dealing with a home care population that frequently has multiple diagnoses—the choice of a primary diagnosis to report could be unduly influenced by payment incentives. If the case-mix system were to consider multiple diagnoses simultaneously, the problem of incentive impacts on reporting might be reduced, but at the expense of more complexity in the adjuster. High predictive accuracy could outweigh these disadvantages, but the R-squared of the diagnostic model was not appreciably higher than the simpler clinical model.

The case-mix project analytic work occurred in three stages: early exploratory analyses, clinically driven development work, and refinements.

Early data analyses. We began exploratory analyses with the 4,303 observations available early in the analysis phase. These analyses relied mostly on regression equations to begin to understand which OASIS and other assessment variables might play an important role in an eventual case-mix adjuster, and to gauge how much variation in resource use beyond case mix alone could be explained in a mathematical model that included factors such as agency characteristics, economic characteristics in the agency's environment, and events taking place during the home health visit. These exploratory regressions suggested that up to .47 of the variation in resource use could be explained using regression analyses that accounted for a range of causal factors encompassing more than case mix. The equations included variables to measure clinical, functional, home environment, agency, and economic factors; home health treatment variables; and unusually time-consuming events taking place during visits. These analyses highlighted several potentially appropriate and powerful variables in the data, such as preadmission location of the patient; certain acute conditions (orthopedic, neurologic, open wounds and lesions, diabetes); the presence of an ostomy; and functional dependence in locomotion. These models further suggested that restricting the explanatory variables to a subset of purely clinical and functional patient characteristics alone would produce an R-squared of approximately .20.

Clinically driven case-mix models: The project's goal from the outset was to develop a case-mix adjuster that defines a number of mutually exclusive patient groups that could be associated with differing resource use. Another criterion for the grouping system is that it should be clinically meaningful to the home health clinicians using it, by making use of recognized clinical categories and by being consistent with the clinical diagnostic process. A further criterion was simplicity; ideally, the system should be comprised of a limited number of mutually exclusive groups, and rules for classifying patients into groups should be straightforward.

As described in their project report (Abt Associates, Inc., Second Interim Report, August 1999), these objectives were approached by the Abt Associates nurse-clinicians through a combination of professional experience and study of previous work in the field reported in the literature. They first focused on identifying clinically significant indicators that address patient care needs from the perspective of the home health clinician. To help identify indicators, they considered the following questions: What level of complexity, severity and instability characterizes the patient's clinical condition? How much and what type of assistance does the patient need with activities of daily living? Does the patient require special therapies or high-tech services? What cognitive impairments, behavioral characteristics, risk factors, and environmental conditions affect the amount and type of care this patient will require? The Abt team then proceeded to review the patient assessment variables as a source of information for the indicators. The resulting list of variables was reviewed in light of several issues:

Policy implications: Some patient characteristics are not suitable as a basis for payment because they raise issues of equity or are otherwise questionable from a policy perspective. For example, the assessment's race and education variables were excluded, as were measures of the patient's social or physical environment (for example, unsanitary or unsafe conditions). Similarly, a case-mix adjustment system should not discourage assistance from family members of home care patients. Although many observers assume that the availability or efficacy of a caregiver is a significant influence on HHA resource consumption, adjusting payment in accordance with caregiver variables does not seem advisable.

Administrative ease: Initially, the list of assessment items capturing clinically significant indicators included some that were supplemental to the OASIS itself. Incorporating these items in the assessment would require modification of the OASIS data collection procedures and complicate the startup phase for OASIS data collection. We carefully examined the explanatory power of the individual items and sought substitutes for them whenever possible from among the existing OASIS items. We were able to find substitutes for almost all of them with little impact on the explanatory power of the model. The only notable exception was an assessment item about a history of falls, which analysis suggests could raise the explanatory power of the model by about one one-hundredth. However, because this was the only remaining variable that was not obtainable from the existing OASIS collection procedure, we weighed its utility against possible delays and confusion in OASIS implementation and decided not to use it. A utilization variable pertaining to inpatient stays occurring during the home health episode was also seriously considered but ultimately dropped because data limitations prevented us from clearly understanding its impact and because it posed an added data collection burden for home health providers. This item would have required the HHA to report whether a Medicare-covered inpatient stay occurred during the 60-day episode and the length of the stay. This information would be used to determine any adjustment to the case-mix group assignment at the end of the episode.

Other criteria: Reliability-related concerns were also a part of the item selection process. If case-mix variables address characteristics that appear subject to varying interpretation by assessing clinicians, the system could be vulnerable to manipulation by providers or patients. When payment increments are at stake, great care must be taken before accepting items even if they have been proved reliable in other circumstances, such as quality assurance research. For example, items on rehabilitative prognosis and overall prognosis were eliminated on these grounds. Some symptoms may be very short-lived, but if they are present at the time of the assessment they would have an impact on the case-mix adjuster if included. An example is a supplemental item such as ''In last 3 days, noticeable decrease in the amount of food client usually eats or fluids usually consumed?'' We determined that basing payment adjustments on potentially transient signs and symptoms captured by these items is ill-advised because their impact on care delivery is uncertain at best. In addition, diagnoses that were candidates for inclusion in broader diagnosis groups were reviewed by a member of our clinical staff from the perspective of their reliability as markers for resource-intensive conditions.

Incentive effects: Unintended incentive effects could result from using variables that reward providers for negative practice patterns, such as the use of a urinary catheter absent clinical need for the device.

Structure of the system for case-mix measurement. In addition to studying individual variables from the perspectives of explanatory power, policy and administrative implications, and reliability, it was necessary to define the system's decision logic, or structure. Examples of other grouping models developed for research purposes, case-mix classification, risk adjustment or care and treatment were studied to suggest ways of categorizing functional impairment, clinical severity, and other patient characteristics—such as whether to group patient characteristics via distinct dimensions of health status (for example, functional versus clinical); whether to consider bifurcations of groups for which partitioning would produce clinical and statistical meaning (that is, ADL ''splits,'' as the RUG–III system uses); the desirability of symmetrical versus asymmetrical models; and whether to create an indexing system or a categorical system. For example, when considering issues such as cognition, we considered whether these variables would be more appropriately captured within a clinical or functional domain, or whether they would provide more clinical meaning (or statistical power) if used as a binary split (that is, yes/no cognitive impairment) after clinical and functional groups were established.

Similarly, in our consideration of existing classification systems, we examined the clinical value of different structural and operational features of systems. The Nursing Severity Index, for example, adds points per each qualifying nursing diagnosis and sums to a total score. The total score, or index, reflects the patient's severity, with a total index of 34 reflecting the highest severity of illness. Unlike the NSI, the RUG–III classification system is a hierarchical system, with seven general categories that are placed in general order of costs associated with caring for residents. The first category, or top split, is rehabilitation; the last is reduced physical function. As we reviewed these systems, we gave consideration to which type of system seemed least complex for use by home health clinicians, most clinically-intuitive, and most feasible to operationalize, given the nature of the assessment data set.

Abt Associates used a computer package called PC-Group, which creates decision trees whose terminal nodes may be regarded as case-mix groups. This package allows the analyst to ''grow'' the tree interactively, which means considerable judgment can be imposed in selecting and dividing nodes as the tree is constructed.

To produce a workable product with the package, it was necessary for the Abt analysts to summarize their variables first. Based on the conceptual work and literature review conducted during the project, they arrived at a small set of dimensions for summarizing assessment elements. There are separate dimensions for clinical severity; functional status; and service utilization. This organizing principle suggests that patients can be classified along each dimension, and this classification is correlated with resource consumption in home care. In an effort to maximize the clinical utility and explanatory power of the patient classification model, the project team experimented with many variations of each dimension, adding and removing items and examining their effect on the way the models functioned.

The Clinical Severity Dimension. The clinical severity in the final model incorporates three diagnostic categories: Neurologic, Orthopedic, and Diabetes. Specific diagnoses comprising each group were reviewed to ensure that diagnoses used on highly heterogeneous groups of patients would not be included. Inclusion of these diagnoses could weaken the predictive power of the case-mix adjuster. The diagnoses in each group are shown in Table 9. The diagnosis code comes from OASIS item number M230. The clinical dimension also includes the following OASIS items as indicators of clinical severity: status of wounds and ulcers (M0460, M0476, M0488); vision status (M0390); pain frequency (M0420); presence of a bowel ostomy; (M0550) use of parenteral and enteral nutrition, and intravenous therapy or infusion therapy (M0250); dyspnea (M0490); urinary and bowel incontinence (M0530, M0540); and behavioral problems (M0610).

Early versions of the clinical model did not include measures of cognitive, sensory and behavioral impairment which might affect resource use, primarily because statistical analysis did not suggest they were useful in explaining variation. Based upon subsequent review, we determined this was a serious omission from the model, so we renewed attempts to integrate cognition and related indicators into the model. An additional dimension consisting solely of the OASIS neurological, cognitive, sensory, and behavioral (NCSB) variables was created, which produced a minor variance reduction in the overall sample of only .015. Furthermore, the highest degree of cognitive impairment was not consistently related to the highest mean costs.

Since increasing levels of severity of the NCSB variables as a group are not consistently associated with increased resource use, we did not attempt to use them as an independent dimension. Using data from regression analysis, however, we were able to integrate M0390 (vision) and M0610 (behaviors) into the Clinical Severity dimension in a way that did not produce counter-intuitive cost groupings.

Further technical discussion of the statistical results on each variable is found in Abt Associates, Second Interim Report, August 1999, Chapter 3.

The Functional Status dimension. As in the development of the clinical severity dimension, we began by selecting assessment items considered to be potential predictors of increased resource use, focusing on the extent of assistance the patient required with activities of daily living. Early exploration with the available functional indicators suggested OASIS items were equivalent in explanatory power to the supplemental items we tested. We tested restricting the ADLs to late loss ADLs (that is, those ADLs likely to be lost late in life: eating, transferring, toileting, and bed mobility) to see whether the restricted list better predicted resource use in the home-bound elderly, as is the case among the elderly which reside in nursing homes (Williams, Brent C., Brant E. Fries, and William J. Foley, ''Activities of Daily Living and Costs in Nursing Homes,'' *Health Care Financing Review*, 15 (4):117–134 (Summer 1994)). This was not supported. We also experimented with cognition-related variables, based on findings in the literature (Torres, H. A., L. Fratiglioni, Z. Guo, M. Viitanen, E. von Strauss, and B. Winblad, ''Dementia is the Major Cause of Functional Dependence in the Elderly: 3-Year Follow-up Data from a Population-based Study,'' *American Journal of Public Health*, 88:1452–1456 (1998).

In the version of the dimension ultimately used in the Clinical model, ambulation locomotion was integrated and both early-loss and late-loss ADLs were included (while cognitive factors were incorporated into the Clinical Dimension). We dropped the eating and grooming ADLs because they were statistically redundant when the other items (dressing (M0650, M0660), bathing (M0670), toileting (M0680), transferring (M0690), and locomotion (M0700)) were included. M0650 (Dressing Upper body) and M0660 (Dressing lower body) were found to have a significant degree of interaction and therefore were combined. Additional experimentation with the functional status dimension involved testing different schemes for ordering the variables and partitioning subgroups of patients in accordance with measurements on the variables.

None of the variables in the Functional Status Dimension was eliminated due to reliability-related or incentive concerns. Some home health clinicians who reviewed the model in October 1998 commented on the potential of functional status items to be manipulated by providers, who would have an incentive to make patients seem as functionally impaired as possible on admission to home care. However, because the functional status items make an important contribution in predicting home health resource use, and because they are integral to clinical decisionmaking for the home care benefit, they were retained. Furthermore, under the planned Outcome-Based Quality Improvement system for home care, beyond the initial assessment, quality assurance monitoring may help counteract any tendency to overstate the functional dependency of patients. We are soliciting suggestions for approaches, new assessment items, procedures, or other mechanisms that might help guard against mismeasurement of functional status items due to payment incentives.

The Service Utilization Dimension

The Service Utilization dimension contains variables related to services the patient received both before and during the episode of home care. To measure utilization before the start of home care, OASIS item M0170 collects information about inpatient discharges during the 14 days before the assessment. In the analysis of costs associated with pre-admission location, we examined how responses to M0170 were related to mean resource cost. It should be noted that a Medicare SNF stay is always preceded by an acute care hospital stay, so if a patient has a long SNF stay (exceeding 14 days) the acute care stay probably would not be measured by this item. A similar censoring of an acute care event may also occur with rehabilitation stays, although there is no Medicare requirement that such stays be preceded by an acute care hospital stay. On the other hand, if both an acute care stay and a SNF or rehabilitation inpatient discharge occurred within the previous 14 days, it seems likely that the SNF stay or rehabilitation stay was relatively short. We found that patients who are admitted to home care directly from the community are on average more resource-intensive for home care providers than patients who were recently discharged from an acute care hospital and had no evidence from M0170 that they used post-acute institutional care. Patients experiencing both a hospital and SNF/rehabilitation stay within the past 14 days are about as resource-intensive as the patients with no pre-admission stay. Finally, patients for whom only a SNF/ rehabilitation hospital stay is observable within the past 14 days are the most expensive. We theorize that they tended to have relatively long SNF or rehabilitation stays of (at least 14 days), which may suggest that the definition of this group using M0170 is a marker for clinically complicated cases with intensive care needs.

The other variable in the service utilization dimension measures home health therapy hours totaling 8 hours or more during the 60-day episode. In developing the patient classification models, we sought to focus on variables that predicted care needed by the patient, as opposed to care furnished by providers. Ideally, we sought a case-mix adjustor that creates as little incentive as possible for providers to enhance revenues by providing unnecessary services. However, including a variable measuring the receipt of a significant amount of home health therapy (physical, occupational, or speech/ language) improved the R-squared of our models by about .20. The RUG–III system for SNF case-mix measurement also includes an indicator for receipt of therapy. An advantage of paying differentially for therapy cases in the case-mix adjuster is that it will help to maintain access to therapy among home health patients who need it. The threshold of 8 hours targets additional payments for home health therapy to patients with a clear need for therapy. We believe this decision rule will motivate home health providers to efficiently plan therapy evaluation visits and therapy delivery for patients who need little or no therapy.

Additional variables were tested for the services utilization dimension. We decided not to use a variable for previous home health utilization in the past 90 days because, under the influence of payment incentives, it carried the potential to encourage readmissions to home care within the 90-day window. The predictive value of the service utilization was lowered by only .0059 as a result. We also tested the value of including inpatient stay events during the episode. This intervening-stay variable modestly improved the total R-squared for the model. However, as discussed above, it may present substantial data collection burdens for providers.

Scoring Patient Variables and Developing Severity Categories

Variables within the clinical and functional dimensions have differing impacts on resource cost. Before the final refinement phase of model development, we assigned a score to each outcome on each variable based on the increase in mean resource cost associated with each outcome. Within each dimension, the sum of scores for the component variables is correlated with resource consumption in home care. This is consistent with our conceptualization of the clinical, functional, and service utilization components as dimensions along which patients can be classified in accordance with their home health resource consumption.

During the refinement phase of model development, we used regression-adjusted mean resource cost to re-examine the scores. The purpose of the regression was to control for all case-mix variables simultaneously to get a more accurate picture of their respective independent contribution to resource use. Having quantified their contribution via the regression, we could derive more accurate scores for the variables. In addition, we looked for results that could signal redundancy among the variables and tested several interaction terms in the regression. (Interaction terms capture potential synergy among variables.) Both the improved scoring and the interaction terms could potentially improve the explanatory power of the case-mix system. The results of the regression analyses changed some of the scoring and resulted in the merging of some items. A few items were eliminated after examining the regressions, which suggested they were redundant.

The next step in model development was to find score intervals along the clinical dimension and the functional dimension that would define patient groups of relative severity along the respective dimension. Whenever possible, we used ''natural breaks'' in the array of scores in the sample to define the intervals. When partitioning the functional dimension scores, we examined the types of dependencies that would be captured in the intervals, particularly at the low and high end of the functional dimension. We determined the number of intervals also in light of the number of groups that would ultimately be created as more intervals are defined. The R-squared does not improve substantially when one or two more breaks are defined, but the number of groups increases greatly, adding to the complexity of the system.

For the clinical dimension, we classified patients into four levels of impact (minimal, low, moderate, and high), and for the functional dimension, five levels of impact (minimal, low, moderate, high, and maximum). The service utilization dimension is actually comprised of categorical variables that partition patients into four groups of increasing impact on resource use. We assigned scores to each of these four groups in accordance with the increasing impact.

Case-mix Groups. Each dimension contains four or five impact levels or intervals (for example, high, moderate, minimum, and low). For every combination of intervals, there is a case-mix group. For example, patients who are high on the clinical dimension, moderate on the functional dimension, and low on the services utilization dimension are grouped together. Since there are four clinical levels, five functional levels, and four service utilization levels, the case-mix system comprises a total of 80 groups. Half of the groups involve patients with therapy use of at least 8 hours.

In the case-mix research sample, the number of patients in each group varies widely, from few or no patients to between 1,000 and 1,500 in several of the groups (unweighted data). The therapy groups comprise a minority of patients in the sample— 15 percent (unweighted). Approximately 30 percent of the sample fell into the minimal clinical level, 30 percent into the low clinical level, 23 percent into the moderate clinical level, and 17 percent into the high clinical level. Approximately 15 percent of the sample fell into the minimal functional level, 30 percent into the low functional level, 36 percent into the moderate functional level, 11 percent into the high functional level, and 7 percent into the maximal functional level.

III. Audited Cost Report Data Sample Methodology

Audited Cost Report Data

Section 1895(b)(1) of the Act requires the prospective payment amount to include all services covered and paid on a reasonable cost basis under the Medicare home health benefit, including medical supplies. Section 1895(b)(3)(A)(i) of the Act requires the computation of a standard prospective payment amount to be initially based on the most recent audited cost report data available to the Secretary. Under section 1895(b)(3)(A)(i) of the Act, the primary data source in developing the cost basis for the 60-day episode payments was the audited cost report sample of HHAs whose cost reporting periods ended in fiscal year 1997 (that is, ended on or after October 1, 1996 through September 30, 1997).

In February 1998, we directed our fiscal intermediaries (FIs) to conduct comprehensive audits of the cost reports submitted by a sample of HHAs whose cost reporting periods ended in FFY 1997. Each FI received a list of agencies to audit and instructions on how to conduct the audits and report the data obtained.

The sample was designed to be representative of the home health industry in several respects: provider-based versus freestanding, census region, urban versus rural location, and large versus small agencies. Because we anticipated that many agencies in the sample would not be audited because their records were unavailable for a variety of reasons or their cost reporting periods were less than 12 months long, the sample size was adjusted upward by 15 to 20 percent to allow for attrition.

To create national HHA PPS rates, each observation in the final data set is weighted to reflect the national Medicare home health payment experience. For example, the estimates will reflect differences across census regions and urban versus rural areas.

Audit Sample Methodology

To meet these objectives, a statistical sample begins with a list of all HHAs that submit cost reports. The list is referred to as a frame. Considerable effort went into the process of developing the frame for HHAs and identifying units to be included. The frame for this sample excludes all HHAs that are incidental providers (too small) or not likely to yield a full year of cost reporting for the audit period.

Once a frame was developed, we selected a sample. The sample for the HHAs was selected by choosing samples for each provider type (freestanding not-for-profit, freestanding for-profit, freestanding governmental, and provider-based). The provider types are referred to as strata in sampling terms. The design of the sample took into account the number of providers and the variation in cost and beneficiaries in each stratum. The sample was designed to produce estimates from key elements of the audit data with a reasonable level of precision.

A sample selection assumes the frame is complete and each sampling unit appears once and only once in the frame. Unfortunately, after the sample was drawn and fieldwork begun, we found that this assumption was not strictly true for the governmental units.

The problem arises from the fact that multiple providers, referred to as subunits, report under a single cost report. In some cases, multiple providers' numbers corresponding to a single cost report appear on the frame, while in other cases a provider number is a parent possibly with multiple subunits. We then considered the subunits associated with a single cost report as the appropriate sampling unit because there is no way to accurately distribute costs among subunits. The subunits on the frame associated with a single cost report were identified and the listings of individual subunits were regarded as if the appropriate sampling unit had been included a known number of times on the frame list.

This somewhat changed the sample composition. When the sample was drawn for a stratum so that each unit on the list has the same probability of selection (as among the governmental units), the probability that the multiply-listed unit be included in the sample was higher. The higher probability of representation is in proportion to the number of inclusions on the frame list. This is like a drawing in which an individual enters his name (or his family members' names) multiple times to enhance his (or his family's) odds of winning. When one analyzes data from a sample that is biased by giving a higher probability of selection to some units, these units need to be given smaller weights if the estimates are to correctly represent the population that the frame should have enumerated.

That is, the analysis of the sample data must take into account the sampling probabilities by assigning each sampling unit a weight that is less if the probability of inclusion is higher. Indeed, the sample may include the same subunit multiple times, and we retained the values for each time the unit appears in the sample when the proper weights are used.

For purposes of this example, n equals the number of governmental subunits reporting under a single cost report in the frame. Therefore, a governmental cost report is n-times more likely to appear in the sample, and the weights for each occurrence in the sample are reduced by dividing by n. A description of a similar situation involving a household survey based on samples drawn from children in school is described in Morris H. Hansen, William N. Hurwitz, and William G. Madow, *Sample Survey Methods and Theory,* vol. 1 (NY: Wiley, 1953) 59–65. Because households with large families will have a higher probability of being included in the sample, households with large families will be over-represented in the sample unless some adjustment is made. That adjustment can be done, as we did here, by providing weights in the analysis that give less weight to the households that are more likely to be included in the sample.

From the frame we have known totals for the number of units in the cells. Weights were adjusted so that corresponding totals based on the sample match these known cell totals. Even if all units in the sample were successfully audited, the process described above ensures that correct cell totals are obtained from the analysis.

However, when audits are not obtained as intended and the missed units are not in the sample as intended, the weights must be adjusted so that the sample data reproduce the known totals from the frame for key subgroups or cells. The process assigns a larger weight to audited units in the sample similar (in the same cell) to those missed. In the case of the HHA, the cells were defined by the urban or rural area; the four census regions of Northeast, Midwest, South, and West; and provider type. Therefore, the weights were adjusted for the missed sample units to ensure that the units obtained most closely represent the missed units cell by cell.

Summary of the Missing Audits in the Home Health Audit Sample and Results Used to Develop Weights for the Sample

In the home health audit sample design we assumed there would be nonresponse or missing audits for a variety of reasons. The reasons included situations such as the following: the provider no longer existed in order to do the audit, the provider was under